Name: ______________________________

Statistical Inference Worksheet

MA 220-02, Spring 2020

15 April 2020

Write your solutions in the spaces provided. If you run out of room for an answer, continue on the back of the page, and indicate this on the front of the page.

Show all work and explain your reasoning. Use R where appropriate.

Your goal should be to complete **each** problem in **under 15 minutes**. You will complete any problems remaining after the class session for homework.

When you complete a problem, flag me into your breakout room so I can review your solution. If I am currently working with another student, move on to the next problem, then get my attention when I am done working with that student.

1. According to the norms established for a reading comprehension test, eighth graders should average 84.3 on a standardized test. If 45 randomly selected eighth graders from a certain school district averaged 87.8, and the standard deviation of their scores was 10.1, test the claim that the students exceeded the standard at the 0.01 significance level.

   (a) State the claim in terms of an equality / inequality involving the parameter(s) of the population(s).

   (b) State the null and alternative hypotheses associated with the claim.

   (c) Determine a relevant test statistic for investigating the claim. What sampling distribution will you assume for the test statistic? What must be true about the sample and / or population for this sampling distribution to be correct?

   (d) Test the claim using a confidence interval / bound.

   (e) Test the claim using a rejection region.

   (f) Test the claim using a P-value.

   (g) State your conclusion as it relates to the original claim, and interpret the confidence interval / bound in the context of the original claim.

2. In a random survey of 1,000 households in the United States, it is found that 29 percent of the households contained at least one member with a college degree. Does this finding refute the statement that the proportion of all such U.S. households is at least 35 percent? Test the claim at the 0.025 significance level.

   (a) State the claim in terms of an equality / inequality involving the parameter(s) of the population(s).

   (b) State the null and alternative hypotheses associated with the claim.

   (c) Determine a relevant test statistic for investigating the claim. What sampling distribution will you assume for the test statistic? What must be true about the sample and / or population for this sampling distribution to be correct?

   (d) Test the claim using a confidence interval / bound.

   (e) Test the claim using a rejection region.

   (f) Test the claim using a P-value.

   (g) State your conclusion as it relates to the original claim, and interpret the confidence interval / bound in the context of the original claim.

3. A study of the number of business lunches that executives in the insurance and banking industries claim as deductible expenses per month was based on random samples and yielded the following results:

Insurance:	$m = 40$	$\bar{x} = 9.1$	$s_1 = 1.9$
Banking:	$n = 50$	$\bar{y} = 8.0$	$s_2 = 2.1$

Test the claim that executives in the banking industry claim fewer business lunches than executives in the insurance industry at the 0.05 significance level.

(a) State the claim in terms of an equality / inequality involving the parameter(s) of the population(s).

(b) State the null and alternative hypotheses associated with the claim.

(c) Determine a relevant test statistic for investigating the claim. What sampling distribution will you assume for the test statistic? What must be true about the sample and / or population for this sampling distribution to be correct?

(d) Test the claim using a confidence interval / bound.

(e) Test the claim using a rejection region.

(f) Test the claim using a P-value.

(g) State your conclusion as it relates to the original claim, and interpret the confidence interval / bound in the context of the original claim.

4. It has been claimed that more than 40 percent of all shoppers can identify a highly advertised trademark. In a random sample of shoppers, 44 of 100 shoppers were able to identify the trademark. Test the claim at the 0.001 level of significance.

   (a) State the claim in terms of an equality / inequality involving the parameter(s) of the population(s).

   (b) State the null and alternative hypotheses associated with the claim.

   (c) Determine a relevant test statistic for investigating the claim. What sampling distribution will you assume for the test statistic? What must be true about the sample and / or population for this sampling distribution to be correct?

   (d) Test the claim using a confidence interval / bound.

   (e) Test the claim using a rejection region.

   (f) Test the claim using a P-value.

   (g) State your conclusion as it relates to the original claim, and interpret the confidence interval / bound in the context of the original claim.

5. Sample surveys conducted in a large county in a certain year and again 20 years later showed that originally the average height of 400 ten-year-old boys was 53.8 inches with a standard deviation of 2.4 inches, whereas 20 years later the average height of 500 ten-year-old boys was 54.5 inches with a standard deviation of 2.5 inches. Test the claim that the average heights of ten year old boys has increased by half an inch over this 20 year period at the 0.1 significance level.

   (a) State the claim in terms of an equality / inequality involving the parameter(s) of the population(s).

   (b) State the null and alternative hypotheses associated with the claim.

   (c) Determine a relevant test statistic for investigating the claim. What sampling distribution will you assume for the test statistic? What must be true about the sample and / or population for this sampling distribution to be correct?

   (d) Test the claim using a confidence interval / bound.

   (e) Test the claim using a rejection region.

   (f) Test the claim using a P-value.

   (g) State your conclusion as it relates to the original claim, and interpret the confidence interval / bound in the context of the original claim.

6. An airline wants to test the null hypothesis that 60 percent of its passengers object to smoking inside the plane. Explain under what conditions they would be committing a type I error and under what conditions they would be committing a type II error **in the context of this problem**, e.g. **do not** just say, "A Type I Error is rejecting the null hypothesis when the null hypothesis is true."

7. A random sample of size 64 is to be used to test the null hypothesis that for a certain age group the mean score on an achievement test (the mean of a normal population with $\sigma^2 = 256$) is less than or equal to 40.0 against the alternative that it is greater than 40.0. If the null hypothesis is to be rejected if and only if the mean of the random sample exceeds 43.5.

   (a) Find the probabilities of type I errors when $\mu = 37.0, 38.0, 39.0,$ and 40.0.

   (b) Find the probabilities of type II errors when $\mu = 41.0, 42.0, 43.0, 44.0, 45.0, 46.0, 47.0,$ and 48.0.

   (c) Determine the sample size necessary to achieve a power of 0.8 for a significance level 0.01 test of the relevant hypotheses when we assume the true mean is $\mu = 42$.

8. Consider a random sample $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$ from a Gaussian population where σ is **unknown**. We wish to test the following hypothesis:

$$\begin{aligned} H_0 : &\quad \mu \geq \mu_0 \\ H_a : &\quad \mu < \mu_0 \end{aligned}$$

at significance level α.

Test Statistic Method:

1. Determine an appropriate test statistic for this hypothesis test.
2. Determine the rejection region for the appropriate test statistic for this hypothesis test.

Confidence Interval Method:

3. Determine the confidence bound that can be used to test this hypothesis. Denote the true value of μ by μ_0.
4. Determine the condition for rejecting the null hypothesis using the confidence bound.

Equivalence of Two Methods:

5. Show that the interval of values of $\bar{x}$ for which the Test Statistic Method rejects the null hypothesis is equivalent to the interval of values for which the Confidence Interval Method rejects the null hypothesis.

- **Hint:** For each of the two methods, write the interval of values for which the null hypothesis is rejected as an inequality involving μ_0 and $\bar{x}$. Then show that the two inequalities specify the same interval of values for $\bar{x}$.